CLAIMS

What is claimed is:

1.   A system that facilitates modeling unobserved speech dynamics comprising:

an input component that receives acoustic data;

a model component that models speech based, at least in part, upon the acoustic data, the model component comprising model parameters which characterize aspects of the unobserved dynamics in speech articulation, and, which characterize a mapping relationship from the unobserved dynamic variables to observed speech acoustics, the model parameters modified based, at least in part, upon a variational learning technique, and a technique for decoding an underlying unobserved phone sequence of speech based, at least in part, upon a variational learning technique.

2.   The system of claim 1, modification of at least one of the model parameters being based upon a variational expectation maximization algorithm having an E-step and M-step.

3.   The system of claim 2, modification of at least one of the model parameters being based, at least in part, upon a mixture of Gaussian (MOG) posteriors based on a variational technique.

4.   The system of claim 3, the model component being based, at least in part, upon:

$$q(s_{1:N}, x_{1:N}) = \prod_n q(x_n \mid s_n) q(s_n),$$

where x is a state of the model,

s is a phone index,

n is a frame number, and,

q is a probability approximation.

MS302099.1 at top.

State equation: $x_n = A_s x_{n-1} + (I - A_s)u_s + w$,

Observation: $y_n = C_s x_n + c_s + v$,

Probability: $p(s_n = s | s_{n-1} = s') = \pi_{s's}$,

$p(x_n | s_n = s, x_{n-1}) = N(x_n | A_s x_{n-1} + a_s, B_s)$,

5.   The system of claim 2, modification of at least one of the model parameters being based, at least in part, upon a mixture of hidden Markov model (HMM) posteriors based on a variational technique.

6.   The system of claim 1, the model component selecting an approximate posterior distribution relating to the acoustic data and optimizing a posterior distribution by minimizing a Kullback-Liebler (KB) distance thereof to an exact posterior distribution.

7.   The system of claim 1, the model component being based, at least in part, upon a hidden dynamic model in the form of segmental switching state space model.

8.   The system of claim 7, the model component being based, at least in part, upon a switching state-space model for speech applications.

9.   The system of claim 7, the model component employing, at least in part, the state equation:

$$x_n = A_s x_{n-1} + (I - A_s)u_s + w,$$

and the observation equation:

$$y_n = C_s x_n + c_s + v,$$

where n is a frame number,

s is a phone index,

x is the hidden dynamics,

y is an acoustic feature vector,

v is Gaussian white noise,

w is Gaussian white noise and,

C and c are the parameters for mapping from x to y.

10.   The system of claim 7, the model component being expressed, at least in part, in terms of probability distributions:

$$p(s_n = s | s_{n-1} = s') = \pi_{s's},$$

$$p(x_n | s_n = s, x_{n-1}) = N(x_n | A_s x_{n-1} + a_s, B_s),$$

$$p(y_n \mid s_n = s, x_n) = N(y_n \mid C_s x_n + c_s, D_s),$$

where $\pi_{s's}$ is a phone transition probability matrix, $\mathbf{a}_s = (\mathbf{I} - \mathbf{A}_x)\mathbf{u}_s$,

N denotes a Gaussian distribution with mean and precision matrix as the parameters,

C and c are the parameters for mapping from x to y, and,

D represents the covariance matrix of the residual vector after the mapping.

11.   A speech recognition system employing the system of claim 1.

12.   A method that facilitates modeling speech dynamics comprising:

recovering speech from acoustic data based, at least in part, upon a speech model having at least two sets of parameters, a first set of parameters describing unobserved speech dynamics and a second set of parameters describing a relationship between the unobserved speech dynamic vector and an observed acoustic feature vector;

calculating a posterior distribution based on the above model parameters; and,

modifying at least one of the model parameters based, at least in part, upon the calculated posterior distribution.

13.   The method of claim 12 further comprising receiving the acoustic data.

14.   A method that facilitates modeling speech dynamics comprising:

recovering speech from acoustic data based, at least in part, upon a speech model in the form of segmental switching state space model

calculating an approximation of a posterior distribution based on model parameters, the model parameters and the approximation based upon a mixture of Gaussians; and,

modifying at least one of the model parameters based, at least in part, upon the calculated approximated posterior distribution and minimization of a Kullback-Liebler distance of the approximation from an exact posterior distribution.

15.   The method of claim 14 further comprising receiving the acoustic data.

16. The method of claim 14, calculation of the approximation of the posterior distribution being based, at least in part, upon:

$$q(s_{1:N}, x_{1:N}) = \prod_n q(x_n \mid s_n) q(s_n),$$

where x is a state of the model,

s is a phone index,

n is a frame number, and,

q is a posterior probability approximation.

17. A method that facilitates modeling speech dynamics comprising:

recovering speech from acoustic data based, at least in part, upon a speech model;

calculating an approximation of a posterior distribution based on model parameters, the model parameters and the approximation based upon a hidden Markov model posteriors; and,

modifying at least one of the model parameters based, at least in part, upon the calculated approximated posterior distribution and minimization of a Kullback-Liebler distance of the approximation from an exact posterior distribution.

18. The method of claim 17, calculation of the approximation of the posterior distribution being based, at least in part, upon:

$$q(s_{1:N}, x_{1:N}) = \prod_{n=1}^{N} q(x_n \mid s_n) \cdot \prod_{n=2}^{N} q(s_n \mid s_{n-1}) \cdot q(s_1),$$

where x is a state of the model,

s is a phone index,

n is a frame number, and,

q is a posterior probability approximation.

19. A data packet transmitted between two or more computer components that facilitates modeling of speech dynamics, the data packet comprising:

data associated with recovered speech, the recovered speech being based, at least in part, a speech model based upon acoustic data and model parameters, and the model

parameters including at least one articulation parameter and at least one duration parameter.

20. A computer readable medium storing computer executable components of a system that facilitates modeling speech dynamics comprising:

an input component that receives acoustic data; and,

a model component that models speech based, at least in part, upon the acoustic data, the model component comprising model parameters including at least two sets of parameters, a first set of parameters that describe unobserved speech dynamics and a second set of parameters that describe a relationship between the unobserved speech dynamic vector and an observed acoustic feature vector, and, the model parameters modified based, at least in part, upon a variational learning technique.

21. A system that facilitates modeling speech dynamics comprising:

means for receiving acoustic data; and,

means for modeling speech based, at least in part, upon the acoustic data, the means for modeling speech employing model parameters, the model parameters modified based, at least in part, upon a variational learning technique.